



Data mining through neural networks

Himanshu Rehani, Annu

Department of Computer Science and Engineering, Dronacharya College of Engineering, Khentawas, Farukhnagar, Gurgaon, India

Publication History

Received: 13 September 2013

Accepted: 22 October 2013

Published: 1 November 2013

Citation

Himanshu Rehani, Annu. Data mining through neural networks. *Discovery*, 2013, 7(18), 19-22

Publication License



© The Author(s) 2013. Open Access. This article is licensed under a [Creative Commons Attribution License 4.0 \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

General Note



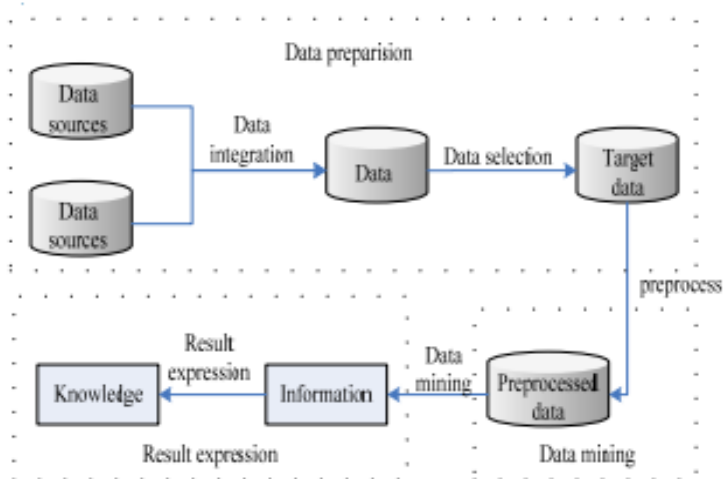
Article is recommended to print as color digital version in recycled paper.

ABSTRACT

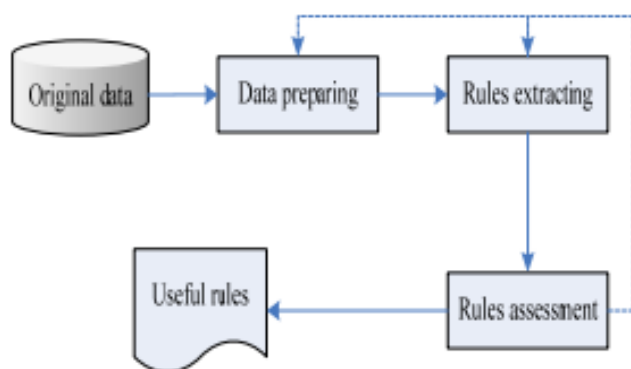
Data mining by using neural networks has become very popular nowadays. As, the neural networks has complex structures, long training time, difficult to understand but they have high acceptance ability for noisy data and high accuracy that are preferable in data mining. In this paper, application of neural network for data mining is researched in detail and the different ways for data mining are also researched.

1. INTRODUCTION

With the rapid development of database technology and the evolution of large number of applications based on database management system, a large volume of data can be stored and the important information in the stored data is hidden. If the information can be retrieved from the database easily, it will create a lot of profit to the companies and the technology of retrieving particular information from a large amount of data stored in database is known as data mining. Generally, data mining is a computational process that involves methods of artificial intelligence, machine learning, statistics and database systems for discovering patterns in large data sets. The overall aim of data mining is to retrieve information from data set and transform it into easily understandable structure. Data mining tools can outlook the future trends and activities to support decision of people. Data mining tools can analyze the whole database of an organisation and can answer the questions like "From which country each employee belongs to?" and other similar problems. Some data mining tools can resolve traditional problems which consumed a lot of time, because they can effectively and efficiently browse the entire database and find some useful information that experts unnoticed.

**Figure 1**

General data mining process

**Figure 2**

Data mining process based on neural network

2. ARTIFICIAL NEURAL NETWORK

Artificial Neural Network (ANN) is a parallel processing network that is capable of machine learning and pattern recognition. They are represented as an interconnected system of neurons that can compute values from inputs by feeding information through network. It is generated with simulating the image intuitive thinking of human on the basis of the research of biological neural network, according to the features of biological neurons and neural network and by simplifying, summarizing and refining. The knowledge of input and output can be expressed easily as it uses the method of parallel processing, non-linear mapping and the structure of neural network itself. The neural network has large complex structures and long training time due to which process of data mining for the applications of neural network was not optimistic. On the other hand, its advantages such as low error rate, high affordability to noise data network pruning algorithms and rules extracting algorithm, makes the application of neural network in data mining favoured by vast of the majority of users.

Structure of Neural Network

A neural network consists of several neurons in different layers. The different layers are input layer, hidden layer and output layer. The "input layer" is the layer where initial stimulus is received. These neurons are connected to a layer of "hidden" neurons which again are connected to either "hidden layer" or "output layer". There can be any number of hidden layers between input and the output

layer but the number of in between layers is limited. Neurons of the different layers are connected to each other in adjoining layers.

Neural Network Topologies

1. Feed-forward neural network: In this type of network, the movement of information is unidirectional i.e. from input node through the hidden node and to the output node. The data processing can extend over a number of layers but no feedback is present. It is specially used in the areas of prediction and pattern recognition.
2. Feed-back neural network: In this type of network, the movement of information is bidirectional. These networks can propagate data from later stages to earlier stages. Feed-back is present in these networks that is the output of a neuron is fed back to input neuron. It is mainly used for associative memory and optimization calculation.

3. DATA MINING ON THE BASIS OF NEURAL NETWORK

As, neural networks are non-linear statistical data modelling tools. Thus, complex relationships between inputs and outputs can be developed using these modelling tools. Finding of patterns can also be implemented using data modelling tools. Neural networks consists of three things i.e the architecture, the learning algorithm and the activation functions. These networks are trained to recognize and retrieve patterns, to solve optimization problems and ill-defined problems. The two main abilities pattern recognition and function estimation makes artificial neural networks so important entity in data mining. The need of automated processing increases due to increase in the size of data sets.

Data mining reaches to the depth of databases. The task for the data mining is classified into two categories:

1. Descriptive data mining

2. Predictive data mining

In descriptive data mining, information is provided to the user in prior to understand what is happening inside the data while predictive data mining allows the user submit records with unknown field values, and system will guess unknown values on the basis of previous patterns discovered from database. The categorization of data mining models depends on the task they perform: Classification and Prediction, Clustering, Association rules. Classification and Prediction is predictive model while clustering and association rules are descriptive models. The most common action performed in data mining is classification. It defines patterns that describe the group to which an item belongs. The examining is done on the basis of existing items that has already been classified and a set of inferring rules. Clustering, another data mining model is similar to classification. The difference between both is that no groups are predefined in the clustering. Prediction is the construction and use of model for accessing the class of an unlabeled object. Another application is forecasting. It is different from predictions as it estimates the future value of continuous variables based on patterns. From last few years, the applications of neural networks to financial forecasting have been popular because of the property of the neural network to mine valuable information from a mass of information and can be used efficiently in financial areas. It has also been proved by different researchers that neural networks performed far better than conventional statistical approaches in data mining. Neural network is one of the most important tool in data mining of data warehouses.

4. DATA MINING PROCESS USING NEURAL NETWORK

The process of data mining consists of three main phases: Data preparation, Data mining, Expression and interpretation of results. The figure 1 shows the flowchart of a general data mining process. While the data mining based on neural network consists of three different phases than that of general data mining process that are data preparation, rules extracting and rules assessment, as shown in figure 2. The three different phases of data mining process using neural network are explained below:

4.1. Data Preparation

This phase define and process the mining data to make it specific data mining model. This is the first important step and plays a critical role in the overall data mining process. This phase consists of following four processes:

1. Data Cleaning

The task of this process is to fill the vacancy value, eliminate the noise data and correct the inconsistencies in data.

2. Data option

This process selects the data arrange and row used in the mining.

3. Data pre-processing

This task of this process is to enhanced process the clean data which has been selected.

4. Data Expression

The transformation of data after pre-processing into the form that can be accepted by data mining algorithm based on neural network. Generally, it transforms sign data into numerical data because the data mining on the basis of neural network can handle only numerical data. The simplest approach of transformation is to establish a table with one-to one correspondence between sign data and numerical data. Another complex approach for the transformation is to adopt appropriate hash function to generate a unique numerical data according to the given string. As, there are a number of data types in relational database, but they all can easily be converted into three logical data types that are sign data, discrete numerical data and serial numerical data. The figure 3 shown shows the transformation of data types. By using a symbol table or hash function, the symbol "Apple" in the figure is transformed into corresponding discrete numerical data. Then, this discrete numerical data can be quantified into continuous numerical data and can also be encoded into coded data.

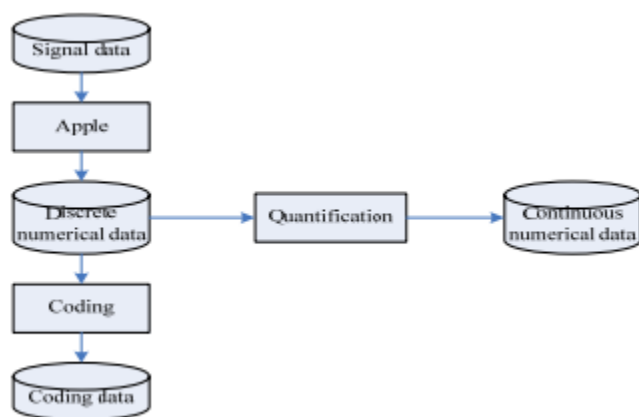
4.2. Rules Extracting

Numerous of methods are present to extract rules, but the most commonly used methods are LRE method, black-box method, the method of extracting fuzzy rules, method of extracting rules from recursive network, algorithm of binary input and output rules extracting, partial rules extracting algorithm and full rules extracting algorithm.

4.3. Rules Assessment

The objective of rules assessment depends on specific application, but in general the rules can be accessed using following objectives:

1. Finding of optimal sequence of extracting rules so that best results can be obtained using the given data set.
2. Test the accuracy of rules extracted.
3. Detect the amount of knowledge that has not been extracted in the neural network.
4. Detect the inconsistencies between trained neural network and extracted rules.

**Figure 3**

Conversion of Data Types

input signal is mapped to the low dimensional space and maintains the input signal in the regional region is space, this is also known as self organisation feature map.

5. TYPES OF DATA MINING

The types of data mining are hundreds but the two most used types are data mining based on self organization and data mining based on fuzzy neural network.

5.1. Data Mining Based On Self Organization

Self-organisation is a process of learning without the help of teachers. The study is done through the important characteristics or some inherent knowledge in group of data. Scholar T. Kohonen considers that the neighbouring modules of the neural network are similar to brain neurons and play different roles. These neurons can be adaptively developed to detect different signal through interaction. These brain neurons are sensitive to different input modes as they play different rules in different parts of the brain space. He also proposed a kind of learning mode in which

5.2. Data Mining Based On Fuzzy Neural Network

As the neural network has strong ability of learning, classification and association, but in process of data mining through neural network, the greatest difficulty is that the output cannot be determined properly. The introduction of fuzzy processing in neural network does not only increase the output expression capacity but also makes the system more stable. The fuzzy networks that mostly use in data mining are fuzzy perception model, fuzzy inference model, and fuzzy ART model.

6. CONCLUSION

At present, data mining through neural network is important area of research and neural network itself very good for solving complex problems of data mining as it possesses characteristics of good robustness, parallel processing, distributed storage and a high degree of fault tolerance. The efficiency of data mining methods can greatly be improved by the combination of data mining methods and neural network model.

REFERENCE

1. Guan Li, Liang Hongjun. Data warehouse and data mining. *Microcomputer Applications*. 1999, 15(9), 17-20
2. Towell G, Shavlik JW. The extraction of refined rules from knowledge-based neural networks [J]. *Machine Learning*, 1993, 13, 71-101
3. Berry JA, Lindoff G. *Data Mining Techniques*, Wiley Computer Publishing, 1997 (ISBN 0-471-17980-9)
4. Zurada JM. *An introduction to artificial neural networks systems*, St. Paul: West Publishing, 1992
5. Fausett Laurene. *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*, Prentice-Hall, New Jersey, USA, 1994